

IKT-SICHERHEITSKONFERENZ 2026

Agentic AI trifft Zero Trust

Wer kontrolliert die KI-Agenten in Ihrem Netzwerk?

Udo Riedel · CTO, DriveLock SE



Der Fahrplan für die nächsten 25 Minuten

01 Zwei Szenarien, ein Kontrollverlust

02 Warum klassische Endpoint-Security versagt

03 Fünf neue Angriffsflächen

04 Zero Trust für KI-Agenten

05 NIS2 und EU AI Act

06 Fazit: eine steile These

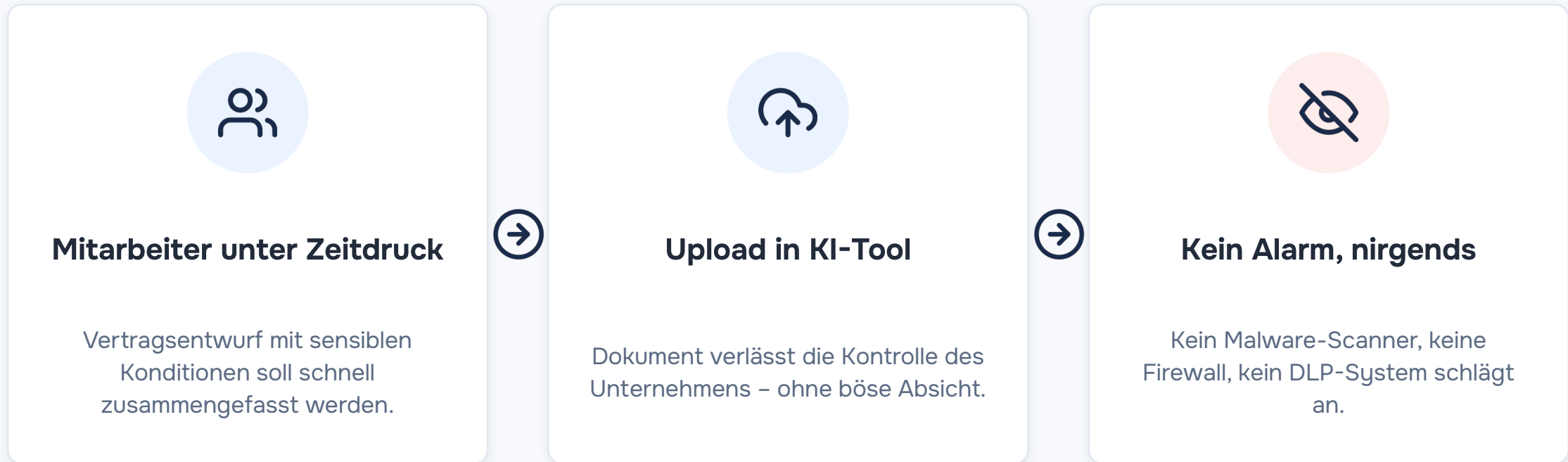
Ein Dokument spricht zum Agenten – nicht zum Menschen



Kein Exploit, kein Schadcode – nur Missbrauch legitimer Rechte durch manipulierte Entscheidungslogik.

SZENARIO 2

Kein Angriff. Nur Bequemlichkeit – mit Folgen.



Theoretisch DLP-abdeckbar. Faktisch denkt kaum jemand daran, KI-Tools als eigene Kategorie zu behandeln.

„Das hier ist keine Zukunftsmusik. Das passiert in Ihrem Netzwerk vermutlich heute Nachmittag.“

Endpoint-Security-Konzepte der letzten zwei Jahrzehnte wurden für Nutzer, Prozesse und Geräte gebaut. KI-Agenten passen in keine dieser Kategorien sauber hinein.

02 / 06

Warum klassische Endpoint-Security versagt

Drei etablierte Entitäten – und eine vierte, die durchs Raster fällt.

Nutzer, Prozess, Gerät – und der Agent



VIERTE ENTITÄT

Handlungslogik entsteht dynamisch aus Modellverhalten, Kontext und Werkzeugen – nicht mehr vollständig im Code verankert.

03 / 06

Fünf neue Angriffsflächen

Kein exotisches Zero-Day-Szenario – Governance-Lücken, die längst existieren.

Fünf Facetten desselben Grundproblems



Prompt-Injection-Rechteausweitung

Manipulierte Eingaben führen zu unautorisierten Aktionen.



Nutzer-initiierte Exfiltration

Vertrauliche Dokumente landen aus Bequemlichkeit in KI-Tools.



Agent-Ketten-Exfiltration

Mehrstufige Datenflüsse, die einzeln unauffällig wirken.



Supply-Chain-Risiko

Drittanbieter-Plugins erweitern die Angriffsfläche unsichtbar.



Fehlende Auditierbarkeit

Blackbox-Problem bei Incident Response und Compliance.

Supply-Chain-Risiko durch Drittanbieter-Agents



Kaum ein Agenten-Einsatz ist vollständig Eigenentwicklung.

Plugins, Frameworks und Werkzeugbibliotheken erweitern die Angriffsfläche – ohne dass dies im eigenen Code sichtbar wird.

Warum das der Kernfall für Application Control ist:

Ohne verbindliche, durchsetzbare Kontrolle darüber, welche Agenten-Komponenten überhaupt ausgeführt werden dürfen, bleibt jede weitere Absicherungsebene Stückwerk. Nachladen zur Laufzeit und Bezug über Marktplätze erschweren klassische Freigabeprozesse zusätzlich.

Das Blackbox-Problem: Wer hat wann was getan?



Welche Instruktion hat der Agent erhalten?



Welche Zwischenschritte hat er gewählt?



Welche Daten gingen an welches Werkzeug?



Für die meisten Umgebungen: keine belastbare Antwort. Klassisches Logging erfasst Ein- und Ausgabe – nicht die Entscheidungsschritte dazwischen.

04 / 06

Zero Trust für KI-Agenten

„Never trust, always verify“ – ausgeweitet auf eine dritte Entität.

Agenten als eigene Identitätsklasse



Nutzer

Etabliert seit Jahren

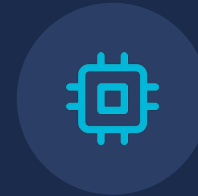
never trust,
always verify



Gerät

Etabliert seit Jahren

never trust,
always verify



Agent

Der fehlende Baustein

never trust,
always verify

Vom Prinzip zur Plattform

01



Application Control

Agenten & Plugins als
kontrollierbare
Ausführungseinheit

02



Policy Engine

Granulare Rechte pro Agent-
Kontext, nicht pauschal

03



Data Protection

DLP für KI-Uploads und
Agent-Ketten

04



Audit & Reporting

Lückenlose
Nachvollziehbarkeit jeder
Aktion

Erst im Zusammenspiel aller vier Ebenen wird die Angriffsfläche strukturell geschlossen – nicht punktuell.

05 / 06

NIS2 und EU AI Act

Der regulatorische Hebel für das, was ohnehin überfällig ist.

Zwei Rechtsrahmen, ein Handlungsdruck



NIS2

Risikomanagement & Nachweispflicht

Betreiber wesentlicher und wichtiger Einrichtungen müssen Sicherheitsvorfälle nachweisbar aufarbeiten können. Ohne Audit-Trail von Agent-Aktionen keine belastbare Nachweisführung.



EU AI Act

Betreiberpflichten, nicht nur Anbieterpflichten

Trifft explizit auch Organisationen, die eigene KI-Systeme betreiben – etwa Behörden, die aus Datenschutzgründen bewusst nicht in die Cloud gehen.

Ihr wolltet aus der Cloud raus, um Kontrolle zu behalten – und genau dafür holt euch jetzt der AI Act ein.

„Wer KI-Agenten wie gewöhnliche Software behandelt, hat den Kontrollverlust bereits akzeptiert.“

Agenten brauchen eine eigene Sicherheitskategorie – mit eigener Identität, eigenen Rechten und eigener Nachvollziehbarkeit.

Vielen Dank!

Fragen & Diskussion

Udo Riedel · CTO, DriveLock SE
udo.riedel@drivelock.com